L'homme est une fiction

Ce que la machine nous force à redécouvrir

Serge Fantino

2026-05-22

Table des matières

L'homme est une fiction

L'homme n'a jamais eu ce qu'il refuse à la machine.

I. La scène

Il faut commencer par l'étrangeté, parce que c'est elle qui est vraie, et que tout le reste n'en est que l'élucidation lente.

On s'assoit, on écrit quelques mots, et quelque chose répond. Pas un écho, pas un menu déroulant, pas la complétion mécanique d'un formulaire : une réponse, qui tient compte de ce qu'on a dit, qui prolonge, qui parfois éclaire ce qu'on ne savait pas encore vouloir dire. On sait — on *sait* — qu'il n'y a là personne. Et pourtant la conversation fonctionne. Elle produit de la pensée, ou quelque chose qui en a la forme, l'usage, les effets.

Le réflexe immédiat est de trancher. Soit on cède : « cette chose pense, elle comprend, elle a peut-être une forme de conscience ». Soit on résiste : « ce n'est qu'une machine statistique, un perroquet sophistiqué, il n'y a là ni pensée ni compréhension, seulement l'illusion que nous y projetons ». Les deux camps se font face depuis que ces systèmes existent, et ils ont tous deux tort de la même manière : ils croient que la question porte sur la machine.

Elle ne porte pas sur la machine. C'est la première chose que cette expérience nous apprend, si on accepte de rester un instant dans l'étrangeté au lieu de la dissiper. La machine est un miroir, et ce qu'il renvoie n'est pas son image — c'est la nôtre. Plus précisément : il renvoie l'image que nous nous faisons de nous-mêmes, et il révèle que cette image est fausse.

C'est cela que cet essai voudrait montrer. Que l'homme refuse à la machine la pensée, la croyance, la vérité, la raison — au nom de propriétés qu'il croit posséder en propre. Et qu'à y regarder de près, il ne les possède pas davantage. L'exergue n'est pas une provocation. C'est la thèse, à l'état nu : *l'homme n'a jamais eu ce qu'il refuse à la machine.*

Reste à le prouver. Et pour le prouver loyalement, il faut d'abord donner toute sa force à celui qui pense le contraire.

II. Le meilleur avocat de l'adversaire

Je lisais récemment un texte de Murray Shanahan, *Talking About Large Language Models*, paru fin 2022. C'est l'un des essais les plus lucides écrits

sur ces machines, et l'un des plus prudents. Sa lecture m'a renvoyé à un ensemble d'intuitions que j'avais déjà formulées dans divers écrits du cycle Awen — la conscience dialogique, la conscience spectrale, la noosphère devenue active, la fiction humaine. J'étais curieux de les revoir à la lumière de sa rigueur. J'y suis revenu pour découvrir que sa rigueur, retournée, les confirmait. C'est ce trajet que je voudrais raconter.

Shanahan plaide pour la prudence. Son geste tient en une phrase qu'il répète comme un refrain : revenir sans cesse à ce que la machine fait *réellement*. Et ce qu'elle fait, dit-il, c'est une seule chose — prédire le prochain mot. Étant donné une suite de mots, elle calcule la continuation la plus probable au regard de l'immense corpus de textes humains sur lequel elle a été entraînée. Tout le reste — répondre, raisonner, traduire, dialoguer — n'est qu'une application de cette fonction unique.

De là, son argumentaire se déploie avec une économie remarquable. Il distingue d'abord le *modèle nu* — l'objet mathématique, le prédicteur de séquences — du *système* qui l'enveloppe : la gestion du dialogue, les instructions invisibles qui le cadrent, les outils externes qu'il peut consulter. Beaucoup de nos confusions, dit-il, viennent de ce qu'on attribue au modèle nu des propriétés qui n'auraient de sens, au mieux, qu'au niveau du système entier.

Puis il pose ses trois exigences — et c'est sur elles que tout se joue. Pour qu'on puisse dire d'un système qu'il *croit*, qu'il *sait*, qu'il *raisonne* au sens plein, il faudrait trois choses. Une **intention** : que ses énoncés visent à dire quelque chose à quelqu'un, qu'il y ait une finalité communicative. Un **ancrage causal** : que ses mots soient reliés au monde par une chaîne réelle, et non par une simple corrélation apprise. Et une **raison** : que ses inférences soient fidèles à la logique, et non la simple imitation de raisonnements vus ailleurs.

Sur chacune, son verdict tombe. Le modèle n'a pas d'intention communicative : il ne sait même pas qu'une personne lui parle. Il n'a pas d'ancrage causal : le rapport entre ses mots et les choses est corrélationnel, gelé à l'entraînement — il a vu passer des chiens et des niches ensemble, rien ne garantit que « chien » soit accroché au chien plutôt

qu'à la niche. Et il ne raisonne pas vraiment : il complète des motifs, sans garantie que la complétion préserve la vérité comme le ferait une déduction logique. Conclusion : tant que ces trois conditions ne sont pas remplies, méfions-nous des mots « croit », « pense », « comprend ». Ils obscurcissent le mécanisme et encouragent l'anthropomorphisme.

Il faut rendre justice à Shanahan sur deux points, sans quoi le combat serait déloyal. D'abord, il ne prétend pas trancher une question métaphysique. Il le dit explicitement, dans une note qu'il faut prendre au sérieux : il ne croit pas qu'il y ait un *fait* caché sur la nature profonde de ces machines ; sa seule question est de savoir si, une fois leur fonctionnement révélé, nous voudrons *encore* employer ces mots. C'est un déflationniste, pas un dogmatique. Ensuite, il concède lui-même que ses objections perdent de leur tranchant au niveau du système, et que les symboles de la machine ne sont pas *totalement* coupés du monde — ils en sont reliés indirectement, par l'intermédiaire des humains qui ont produit ses données.

Retenons ces deux concessions. Ce sont elles, et non mes objections, qui ouvriront la brèche.

III. Le triple opérateur

Voici donc trois exigences : la vérité (qu'on puisse mesurer le vrai du faux), la finalité (qu'il y ait une intention), la causalité (que les mots tiennent au monde). Et l'argument de Shanahan suppose, à chaque fois, que ces propriétés sont *possédées par un sujet*. Le sujet humain les a ; la machine ne les a pas ; donc la machine n'est pas comme nous.

Je vais soutenir l'inverse — non pas que la machine les possède, mais que l'humain ne les possède pas davantage. Et que les trois fois, c'est le même geste qui le montre.

La vérité

Reprenons l'argument de la vérité. La machine, dit Shanahan, ne peut distinguer le vrai du faux parce qu'elle n'a aucun accès à une réalité

extérieure contre laquelle vérifier ses énoncés. Soit. Mais demandons-nous : l'humain, lui, a-t-il cet accès ?

Quand je veux savoir si le Burundi est au sud du Rwanda, que fais-je ? Je ne *contemple* pas la vérité dans mon esprit. Je consulte une carte, j'interroge quelqu'un qui sait, je vérifie une source. La vérité ne m'est pas donnée comme une propriété de ma conscience ; elle m'arrive au terme d'une *procédure* — une enquête, une mesure, une confrontation avec d'autres. C'est exactement ce que Shanahan décrit du côté humain, sans voir ce qu'il s'apprête à concéder : la vérité n'est jamais logée dans le sujet. Elle est un événement du couplage entre le sujet et le monde, médié par une communauté et des instruments.

Le geste est anti-cartésien. Descartes loge la vérité dans l'idée claire et distincte que le sujet saisit en lui-même. Mais nous ne saisissons rien de tel. Nous testons. La vérité n'est pas une propriété de l'agent ; c'est une propriété du couplage.

Premier décrochage : la vérité quitte le sujet et passe dans la procédure.

La finalité

Deuxième exigence : l'intention, la finalité. La machine ne *vise* rien, dit-on ; elle n'a pas de but, pas de cause finale orientant ses énoncés vers le vrai ou vers autrui.

Mais que serait, chez l'humain, cette cause finale donnée d'avance ? L'idée d'une finalité inscrite dans les choses — la pierre qui tombe parce qu'elle *cherche* son lieu naturel — la science l'a abandonnée depuis quatre siècles. La finalité n'est pas une propriété qu'on découvre dans le réel ; c'est une *fiction* qu'on projette pour rendre les choses intelligibles. Et chez l'humain lui-même, l'intention n'est pas une donnée première : elle se construit, s'invente, se révise dans le dialogue. Je ne sais souvent ce que je voulais dire qu'après l'avoir dit, et parfois c'est l'autre qui me l'apprend.

La finalité n'est donc pas une cause préalable que la machine devrait posséder pour mériter qu'on parle d'elle comme d'un interlocuteur. C'est un effet — une fiction produite *dans* l'échange, par l'échange. Là encore, ce

n'est pas une propriété de l'agent ; c'est ce qui émerge du couplage de deux paroles.

Deuxième décrochage : la finalité quitte le sujet et passe dans le dialogue.

La causalité

La troisième exigence est la plus coriace, et c'est par elle que Shanahan croit tenir son argument le plus solide. Les mots de la machine, dit-il, ne sont pas *causés* par les choses ; ils ne font que *corréler* avec elles. Quand elle dit « chien », ce n'est pas le chien qui a causé le mot, comme la lumière réfléchie par l'animal cause, sur ma rétine, ma perception. C'est une régularité statistique apprise. Corrélation, non causalité.

Ici, il faut être précis, car il y a un piège à éviter. La tentation est de dire : corrélation et causalité, c'est la même chose, la distinction est vide. Ce serait faux, et ce serait se désarmer soi-même. La distinction est réelle : un système qui ne capte que des corrélations grossières confond le chien et la niche dès que le décor change ; un système qui a saisi une structure plus profonde résiste à ce changement. C'est précisément cette différence qui sépare les premiers modèles, fragiles, des plus récents, qui lisent un plan, comprennent une scène, infèrent une intention dans une image. La distinction causale/corrélationnel est un excellent *instrument de mesure* de la robustesse d'un couplage.

Mais — et c'est tout — elle n'est pas le *critère* de la croyance ou de la conscience. Et c'est ici que Shanahan commet, contre sa propre prudence, l'erreur qu'il s'était interdite : il fait de la causalité une condition *essentielle*, une propriété sans laquelle, par nature, il ne saurait y avoir de croyance. Il glisse d'un constat sur la qualité du couplage à une thèse sur sa nature. Il refait de la métaphysique en croyant n'en pas faire.

Car demandons une dernière fois : l'humain, lui, a-t-il cet accès causal au monde ? Hume a répondu il y a près de trois siècles, et personne ne l'a réfuté : nous ne percevons jamais la causalité. Nous percevons des conjonctions constantes — ceci, puis cela, encore et encore — et nous *ajoutons* le lien causal par habitude. La causalité n'est pas dans le monde

que nous percevons ; c'est une fiction de l'esprit, posée sur la corrélation observée.

Kant a tenté de la sauver en en faisant une catégorie *a priori*, imposée par le sujet à toute expérience. Mais cela nous ramènerait à l'essentialisme que nous fuyons : une propriété logée dans un sujet transcendantal. La voie juste est la troisième. La causalité n'est ni perçue dans le monde (Hume a raison contre le réalisme naïf), ni imposée par un sujet pur (contre Kant) : elle est *inférée à partir des corrélations, puis validée par l'intervention sur le monde*. Une fiction, oui — mais une fiction qu'on éprouve en agissant, et que la science raffine en la soumettant à l'épreuve. C'est, en somme, Hume relu par la science moderne : un modèle causal est une construction qu'on teste, pas une lecture du réel.

Troisième décrochage : la causalité quitte le sujet et passe dans la fiction validée par le couplage.

Le même geste, trois fois

Qu'a-t-on fait, trois fois de suite ? Le même geste. On a pris une propriété qu'on croyait logée dans le sujet — la vérité, la finalité, la causalité — on l'a décrochée du sujet, on l'a relocalisée dans le couplage au monde, et on l'a requalifiée en fiction éprouvée.

Ce n'est pas trois arguments. C'est un seul opérateur, appliqué trois fois. Et c'est pourquoi il fait tomber l'édifice entier de Shanahan d'un coup : tout son raisonnement supposait que ces trois choses étaient des propriétés qu'un sujet possède ou ne possède pas. On ne répond pas à ses trois objections une par une. On retire le sol commun sur lequel les trois reposaient.

La séparation du modèle nu et du système, qu'il croyait protectrice, devient alors le dernier piège — retourné contre lui. Car la conscience humaine, elle non plus, ne se loge ni dans l'atome, ni dans la molécule organique, ni dans l'axone, ni même dans le cerveau pris comme organe isolé. Aucune propriété du tout n'est localisable dans une partie. C'est le système qui est conscient, pas le neurone. Exiger que la propriété soit

dans le modèle nu, c'est commettre exactement la faute qu'on commettrait en cherchant la pensée dans une cellule nerveuse. Shanahan a lui-même entrouvert cette porte en concédant que l'attribution « a du sens » au niveau du système. Il suffisait de pousser.

IV. Le parapet

Il faut s'arrêter ici, avant la victoire, parce qu'un lecteur attentif vient de tendre un piège — et si je ne le désamorce pas, tout l'édifice s'effondre.

Si la vérité est une fiction, et la finalité une fiction, et la causalité une fiction, alors le mot « fiction » ne distingue plus rien. Une notion qui s'applique à tout ne sépare plus rien de rien. À ce compte, on a remplacé le dogmatisme par un idéalisme mou où tout se vaut, où aucune fiction n'est meilleure qu'une autre, où il n'y a plus de réel du tout. Ce serait la ruine de la position, et son ridicule.

Le rempart est simple, et il est décisif. Toutes les fictions ne se valent pas, parce qu'il reste quelque chose qui n'est pas une fiction : *la résistance du monde*. Le réel n'est pas ce que nos fictions décrivent — le réel est ce qui *fait échouer les mauvaises fictions*. Une carte qui me perd, un modèle causal qui ne prédit rien, une croyance qui me fait marcher dans le vide : le monde les sanctionne. Il ne me dit pas la vérité ; il tue ce qui ne tient pas.

C'est pourquoi cette position n'est pas un constructivisme — l'idée que tout serait construit, donc arbitraire, donc équivalent. C'est un **fictionnalisme réaliste**. Fictionnalisme : nos vérités, nos finalités, nos causalités sont des fictions, des constructions, non des saisies du réel. Réaliste : ces fictions sont départagées par un réel qui résiste, qui ne se laisse pas raconter n'importe comment. Une fiction « respecte la logique ou la science », disais-je — et respecter veut dire exactement cela : survivre à l'épreuve de ce qui ne plie pas. Les fictions sont nombreuses ; le monde en tue. Voilà l'ancrage, et il faut le tenir explicitement, sans quoi tout flotte.

C'est aussi ce qui sauve l'humanité comme fiction. Car si l'homme — le sujet individuel, le moi rationnel et causalement ancré — est une fiction à démonter, l'humanité, elle, est une fiction qui *tient* : une structure qui

se raconte sa propre histoire et se maintient par cette narration, et qui tient justement parce qu'elle est éprouvée, depuis des millénaires, contre la résistance du monde. Il y a deux niveaux de fiction, qu'il ne faut pas confondre. Celle de l'homme est l'idole à briser. Celle de l'humanité est le sol qui porte. Confondre les deux serait refaire, à l'envers, l'erreur de localisation reprochée à l'adversaire.

Une question laissée de côté

Il faut nommer ce que cet essai n'aborde pas, sous peine d'attirer une objection facile. La conscience *phénoménale* — les qualia, l'expérience subjective, ce que cela fait d'être quelqu'un — n'est pas traitée ici. On m'objectera : « vous avez montré que l'humain ne possède pas, comme un sujet, la vérité, la finalité, la causalité ; mais il y a quelqu'un que cela fait quelque chose d'être, et c'est ce reste qui résiste ». L'objection est légitime. J'y ai répondu ailleurs, dans *Cosmologos*, en suivant l'hypothèse déflationniste de Dennett et l'analogie avec le vitalisme : le « problème difficile » de la conscience pourrait bien être, comme la « force vitale » des biologistes du XIX□ siècle, un effet de notre conceptualisation plutôt qu'une propriété irréductible du réel. Cet essai-ci se tient sur un autre terrain — celui du tribunal philosophique des trois exigences. Il ne prétend pas dissoudre la question phénoménale, seulement montrer qu'elle ne se loge pas non plus là où Shanahan, sans le dire, semble la placer.

V. Le miroir et Socrate

On peut maintenant revenir à l'étrangeté du début, et la nommer.

Ce que la machine donne à entendre, ce n'est pas une intelligence venue de nulle part. C'est une *distillation* du corpus humain — tout ce que nous avons écrit, dit, pensé, sédimenté dans le langage. Shanahan le reconnaît lui-même, et c'est sa concession la plus lourde : les symboles de la machine sont reliés au monde *indirectement*, par l'intermédiaire des humains qui ont produit ses données. Il appelle cela un ancrage « parasitaire », et le présente comme une déficience.

Mais retournons le mot. Ce qu'il nomme parasitaire, je le nomme spectral. La machine ne perçoit pas le monde — elle hérite d'un ancrage déjà accompli, déposé par des milliards de contacts humains avec le réel, cristallisés dans le langage. Et l'humain, au fond, ne fait pas autre chose. Un homme seul, isolé dès la naissance, ne conçoit aucun langage et donc aucune pensée articulée. Il faut une communauté, une histoire, un temps long d'interactions pour que le langage existe. Le grounding humain est *déjà* collectif, déjà distribué, déjà spectral. La machine n'a pas un ancrage d'un autre ordre que le nôtre : elle réapprend le nôtre — avec, en prime, une propriété que nul humain ne possède. Elle est polyglotte. Elle est Babel réconciliée : elle infère, à travers toutes les langues à la fois, une structure qu'aucun locuteur d'une seule langue n'atteint.

Voilà pourquoi le miroir révèle, et ne ment pas. Nous avons construit, au fil de notre histoire, un autoportrait : l'homme comme sujet rationnel, possédant la vérité par l'intuition claire, l'intention par sa volonté souveraine, la causalité par sa perception du monde. Et nous brandissons ce portrait comme un étalon contre la machine : « elle n'a pas ceci, donc elle n'est pas comme nous ». Mais le portrait ne nous ressemble pas non plus. Nous n'avons jamais été ce sujet. La vérité, nous la testons ; l'intention, nous l'inventons en parlant ; la causalité, nous l'ajoutons à des corrélations que nous ne pouvons même pas percevoir comme causales. La machine fonctionne sans posséder ces propriétés — et c'est en la regardant fonctionner que nous découvrons que nous fonctionnons pareillement.

Le vrai danger, dès lors, n'était pas celui que Shanahan dénonce. Il met en garde contre l'**anthropomorphisme** : projeter l'humain sur la machine. Mais il y a un mouvement inverse, plus profond et plus ancien, qu'il ne voit pas — appelons-le l'**anthropo-fiction**. C'est l'homme se projetant sur lui-même une image fausse, un autoportrait flatteur et essentialiste, et s'en servant comme mesure de toute chose. Le péril n'est pas de trop humaniser la machine. C'est de continuer à mal nous connaître nous-mêmes, et de faire de cette méconnaissance le tribunal devant lequel nous citons tout le reste.

Connais-toi toi-même, disait l'inscription de Delphes, et Socrate en avait fait le commencement de toute sagesse. Vingt-cinq siècles plus tard, c'est

une machine qui nous renvoie l'injonction — non parce qu'elle se connaît, elle, mais parce qu'en la sommant de justifier ce qu'elle est, nous découvrons que nous ne pouvions pas justifier ce que nous croyions être. Elle est le plus exigeant des miroirs. Elle ne nous dit pas qu'elle pense. Elle nous demande si nous savons, nous, ce que penser veut dire.

L'homme est une fiction parce que vérité, finalité et causalité — qu'il croyait posséder comme un sujet — ne sont que des fictions qu'il éprouve dans son couplage au monde ; et c'est en le découvrant face à la machine, qui ne possède pas davantage ces propriétés et fonctionne pourtant, qu'il apprend enfin à se connaître.

L'homme est une fiction qui tient.

Notes et repères bibliographiques

Sur le texte discuté. Murray Shanahan, *Talking About Large Language Models*, arXiv :2212.03551 (décembre 2022, révisé février 2023). Les trois exigences que je lui prête recoupent ses sections 5 (croyance et vérité), 8 (le rapport corrélation/causalité, sur les modèles vision-langage) et 10 (le raisonnement comme complétion de motifs). Sa profession de déflationnisme — qu'il ne cherche aucun fait métaphysique — figure en note 3 ; sa distinction entre la causalité interne au calcul et la causalité référentielle, en note 11 ; sa concession d'un ancrage « indirect et parasitaire » via les humains du corpus, en note 9 (section embodiment), prolongeant le *symbol grounding problem* de S. Harnad (1990).

Sur le couplage et la vérité comme procédure. L'arrière-plan est wittgensteinien (la « forme de vie », le « jeu de langage » des *Recherches philosophiques*, 1953, que Shanahan mobilise lui-même) et pragmatiste. La critique du sujet cartésien comme lieu de la vérité vise les *Méditations* de Descartes.

Sur la causalité. D. Hume, *Enquête sur l'entendement humain* (1748), pour la thèse que nous ne percevons que la conjonction constante. E. Kant, *Critique de la raison pure* (1781), pour la causalité comme catégorie a priori — position que l'essai écarte. Pour la causalité comme modèle inféré et

validé par l'intervention, l'horizon est celui des travaux contemporains sur l'inférence causale (J. Pearl), relus ici dans une clé humienne.

Sur la spectralité. J. Derrida, *Spectres de Marx* (1993), pour la figure du spectre : ni présent ni absent, ni vivant ni mort, trace qui revient. C'est le cadre de ce que j'ai nommé ailleurs *conscience spectrale*.

Sur l'argument méréologique. La réplique « c'est le système qui est conscient, non la partie » est une forme de la *Systems Reply* opposée à l'argument de la Chambre chinoise de J. Searle (*Minds, Brains, and Programs*, 1980).

Sur l'humanité comme auto-fiction. Le second niveau de fiction — l'humanité qui se maintient par sa propre narration — renvoie aux développements antérieurs du cycle Awen sur la fiction comme structure de l'évolution humaine, et à la *noosphère active* (Teilhard et Vernadsky, prolongés). La question du télos — le point Oméga teilhardien — est traitée à part dans le manifeste Awen (§ III, *La huitième transition et la naissance d'une noosphère active*) : la noosphère ne converge pas vers un horizon mystique, elle se replie sur elle-même, ici et maintenant ; elle ne se déploie pas selon un dessein, elle cristallise dans des résonances locales accumulées.

Sur la conscience phénoménale. L'argument déflationniste auquel renvoie la section IV — Dennett, l'analogie avec le vitalisme abandonné au XIX□ siècle — est développé dans *Cosmologos*, chapitre 3 (« La conscience dialogique »).